

dCache @ Fermilab

Dmitry Litvintsev

Dmitry Litvintsev, 6th International dCache Workshop, 17-19 April 2012, DESY Zeuthen

Wednesday, April 18, 2012





Science at Fermilab

 Fermilab: Home of the Tevatron and buffalos
Energy frontier: BSM, precision EWK, Higgs, HF, QCD

• CDF, D0 (ramping down), CMS T I

 Intensity Frontier: BSM, neutrino physics, rare decays

• Astrophysics: dark matter search, high energy cosmic rays, galaxies surveys







dCache @ Fermilab

		total online	on tape	version
()	public	0.1 PiB	2.5 PiB	1.9.5-28
Size	CDF	I.5 PiB	IO PiB	1.9.5-22
	CMS	13.5 PiB	I7 PiB	1.9.5-23

http://cdfdca.fnal.gov:2288/ http://fndca.fnal.gov:2288/ http://cmsdcam.fnal.gov:2288/

ape Backend

- 7 x SL8500 tapr robot librar Enstore HSM
- http://www-ccf.fnal.gov/enstore/
- LTO4 and TIOK tape drives
- Total data on tape 38 PB



Total TBytes on tape per month from Enstore 2012-Apr-15 02:23:07







public dCache



Pool Setup





• Experiments buy pool hardware

• Their data directed to their pools by regex match between ugroup store units and file storageclass extracted from layer 4 (storage_group + file_family combination + HSM name)

psu set regex on psu create unit -store ktev.ktev@enstore psu create unit -store des.*@enstore ... psu create ugroup DESSelGrp psu addto ugroup DESSelGrp des.*@enstore psu create ugroup KTeVReadSelGrp psu addto ugroup KTeVReadSelGrp ktev.ktev@enstore

Name	Partition	Preferences				Unit Choung				Dool Crouns
		Read	Write	Cache	P2p	Omit Groups			root Groups	
DES-link	default	20	20	10	10	DESSelGrp	world-net	any-protocol	-	DESPools
xpDbWrite-link	default	20	20	10	10	world-net	any-protocol	ExpDbBackupSelGrp	-	ExpDbWritePools
TeVRead-link	default	20	0	10	10	world-net	any-protocol	KTeVReadSelGrp	-	writePools
TeVWrite-link	default	0	20	0	10	world-net	any-protocol	KTeVReadSelGrp	-	writePools
nosPrdRead-link	default	20	0	10	10	MinosPrdSelGrp	world-net	any-protocol	-	MinosPrdReadPool
nosPrdWrite-link	default	0	20	0	10	MinosPrdSelGrp	world-net	any-protocol	-	writePools
wDataWrite-link	default	10	20	10	0	world-net	MinosDaqSelGrp	any-protocol	-	RawDataWritePool
read-link	default	20	0	10	10	any-store	world-net	any-protocol	-	readPools
write-link	default	0	20	0	10	any-store	world-net	any-protocol	-	writePools
nigridVolatile-link	volatile	20	20	0	10	world-net	FermigridVolatile	any-protocol	-	FermigridVolPools
	A DECEMBER OF STREET, STRE	and the second se	the second se	the second s						









transfers by experiment @ public dCache



Dmitry Litvintsev, 6th International dCache Workshop, 17-19 April 2012, DESY Zeuthen

Wednesday, April 18, 2012





transfers by country









- required to develop chimera aware Enstore client (encp v3_10e)
- to support direct encp required customized DB triggers to fill location information from layer I and layer4 and enstore2chimera migration stored procedure. See

https://srm.fnal.gov/twiki/bin/view/DcacheCorner/ChimeraEnstore

- converted pnfs -> chimera on Feb 23rd, 2012
 - 14718667 files

pnfsDump	6h52m		
SQL import	9h47m		
enstore2chimera	Ih8m		
import of companion	I7m		
md5sum verification	II3h24m		







All planned downtimes included, public dCache uptime is 99.2%







- migrate pnfs -> chimera @ CDF and D0 (May 1st)
- upgrade public dCache and CDF dCache to 1.9.12-x
- provide nfsv4.1 server for IF experiments @ public dCache





public dCache for IF

- Intensity Frontier (IF) experiment require POSIX I/O
- => use dCache as backup storage. Use srmcp to copy data to BlueArc for access via NFS.
- Concerns of long term scalability of this solution
- We offered test dCache install with nfsV4.1 server for IF use case studies.





nfsV4.1 for IF



- SLF6 VM @ Fermicloud
- 2.6.40 kernel

- dCache 1.9.12. I head node, 2 pool nodes.
- pool nodes 2 RAID6 w/ 4x250GB SATA drives



lozone in cluster mode with sequential r/w tests





One 4GB file per client (twice the VM memory), one client per VM

Multiple clients per VM





mdtest





mdtest w/ multiple MPI tasks create/stat/remove 100 directories w/ 100 zero length files (stats performed with read-your-neighbor option to avoid reading locally cached files)



nfsV4.1 for IF



 further tests with real data and experiments workflows are foreseen